Award Number: W81XWH-12-1-0013

TITLE: Whole Exome Analysis of Early Onset Alzheimer's Disease

PRINCIPAL INVESTIGATOR: Margaret Pericak-Vance

CONTRACTING ORGANIZATION: University of Miami
Miami, FL 33136

REPORT DATE: April 2016

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
April 2016	Annual Progress Report	1Mar2015 - 31Mar2016

4. TITLE AND SUBTITLE
Whole Exome Analysis of Early Onset Alzheimer's Disease

5a. CONTRACT NUMBER
W81XWH-12-1-0013

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Margaret A. Pericak-Vance, Ph.D.

E-Mail: mpericak@med.miami.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Miami, 1400 NW 10th Avenue, Miami, FL 33136

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The primary focus toward identification of Alzheimer disease (AD) risk genes over the past five years has been testing the common disease common variant (CDCV) hypothesis through the use of genome-wide association studies (GWAS) in late onset Alzheimer disease (LOAD). While common variation clearly plays a role in AD there is a growing realization that the CDCV hypothesis is unlikely to explain all the genetic effect underlying AD. One alternative hypothesis invokes multiple rare variants (RV) in one or more genes, each with stronger individual effects than CDCV genes. We designed this project to test the rare variant hypothesis in AD by examining those cases with the most severe phenotype as determine by early onset (EOAD, cases with AAO < 60 years). Although there are three known EOAD genes (PS1, PS2 and APP) they account for only ~60-70% of familial EOAD and even less of sporadic EOAD. Thus, the majority of the genetics of EOAD remains unknown. Until now, large extended families with AD in multiple generations were necessary to identify variants of significant effect contributing to AD risk, however, with the advent of new genomic technologies such as high-throughput sequencing technology, small family aggregates and isolated cases, particularly those with an extreme phenotype of the disorder (such as early onset) can be used. Thus, we will utilize whole exome high-throughput sequencing to identify high risk AD variants that we will further characterize with respect to AD. We will examine both Caucasian and Caribbean Hispanic AD populations. Our two pronged approach includes structural characterization at the DNA level (Dr. Pericak-Vance), and analysis of Caribbean Hispanics (Dr. Richard Mayeux). Comparing across populations will be extremely useful. Specifically, high priority RVs identified through the whole exome analysis will be further explored with multiple strategies. We will also genotype the interesting variants in a large sample of late-onset (LOAD) cases to examine their involvement in all AD. We will thus prepare a list of high priority candidates for additional follow-up and functional analysis.

.

15. SUBJECT TERMS

Nothing Listed

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	19	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

Page

Introduction…….. 4

Body…… 5

Key Research Accomplishments…….. 10

Reportable Outcomes…… 10

Conclusion…… 11

Appendices…… 12

INTRODUCTION:

The primary focus in the identification of Alzheimer disease (AD) risk genes has focused on the common disease common variant (CDCV) hypothesis using genome-wide association studies (GWAS) in late onset Alzheimer disease (LOAD). It is clear that common variants play an important role in AD, the CDCV hypothesis can't fully explain the genetic factors underlying AD. As an alternative, recent genetic studies have focused on the identification of multiple rare variants (RV) in one or more genes, each with string effect sizes. To that end, the current study was designed to test the rare variant hypothesis in AD by examining those cases with the most severe phenotype as determine by early onset (EOAD, cases with AAO < 60 years). There are three known EOAD genes – Presenilin 1 (PS1), Presenilin 2 (PS2), and Amyloid precursor protein (APP) – that account for ~60-70% of familial EOAD cases and fewer in sporadic EOAD and, as such, the majority of EOAD genes remain to be identified. To that end, we will utilize whole exome next generation sequencing (NGS) to identify high risk AD genetics variants. We will examine both Caucasian and Caribbean Hispanic AD populations. Our two pronged approach includes structural characterization at the DNA level (Dr. Pericak-Vance), and analysis of Caribbean Hispanics (Dr. Richard Mayeux). Comparing across populations will be extremely useful. Specifically, high priority RVs identified through the whole exome analysis will be further analysis, including bioinformatics and computational analysis, genotyping of variants in a large sample of late-onset (LOAD), as well as, functional characterization using patient-specific induced pluripotent stem cells (iPSCs). Patient specific iPSC derived from EOAD patient samples bearing genetic variants of interest will be developed and differentiated into forebrain neurons which will be characterized for markers of AD pathogenesis, including expression of pathogenic amyloid beta and tau isoforms. These iPSC-derived neurons will be screened for small molecule drugs that can correct the cellular deficits.

BODY:

Genetic analyses

An overview of the analyses can be seen in Figure 1. Initial analysis of the non-Hispanic White whole-exome sequencing (WES) data identified 5 genes with the same rare, potentially damaging nonsynonymous or LOF variant in two or more EOAD cases and evidence for protein-protein interaction with a known EOAD gene. Several other cases have a rare nonsynonymous or LOF, potentially damaging variant in another variant in these 5 genes. The 5 genes implicated are: HSPG2 (interacts with GRN and APP), CLSTN1 (interacts with PSEN1 and APP), DOCK3 (interacts with PSEN1 and PSEN2), PARK2 (interacts with MAPT, PSEN1, and APP), and OGT (interacts with MAPT). Sanger sequencing and follow-up genotyping of these variants was completed. Five variants were confirmed as variants (see Table 6) and underwent genotyping in our cohort of AD cases and controls. While some of the variants were seen in additional cases, the variants were also found in controls greater than 65, suggesting the variants are not high-impact EOAD variants in the follow-up sample.

We also completed analysis of a comparison between the Alzheimer's Disease Genetics Consortium (ADGC) early onset Alzheimer's disease exome chip case-control association study and the WES produced from this project. We first updated the association analysis to include a 5th cohort, bringing the total N of the sample to 1,292 cases and 5,625 controls. Analysis comparing the rare, high consequence (missense, non-frameshift, loss-of-function) variants in the NHW WES dataset to the ADGC exome chip association results was then conducted. Briefly, nine genes are genome-wide significant at a Bonferonni correction for 7,249 genes tested (P=6.90 x 10^{-6}), including *PSD2* (P=6.98 x 10^{-7}), an endocytic gene with 2 rare, missense variants present in two separate NHW EOAD cases (Table 3). Preliminary bioinformatics analyses shows both *PSD2* variants in the WES cases to have high CADD scores of 27.4 and 28.5 (above 15 considered damaging) (Kirchner et al. 2014). Additionally, the gene, which is exclusively expressed in brain according to The Human Protein Atlas, is significantly overexpressed in both neurons and astrocytes according to the database Brain-RNASeq (Figure 2) (Zhang et al. 2014). Two genes with rare, segregating variants in the Hispanic families (PER3 and PCDHB11) were found to be genome-wide significant as well (1.74 x 10^{-7} and 8.92 x 10^{-7}). Additionally, the gene *IL16*, in which WES found 2 NHW cases and 2 Hispanic families with rare, missense variants, had suggestive significance in the exome chip study (P=8.33 x 10^{-4}). Results have been incorporated into a manuscript.

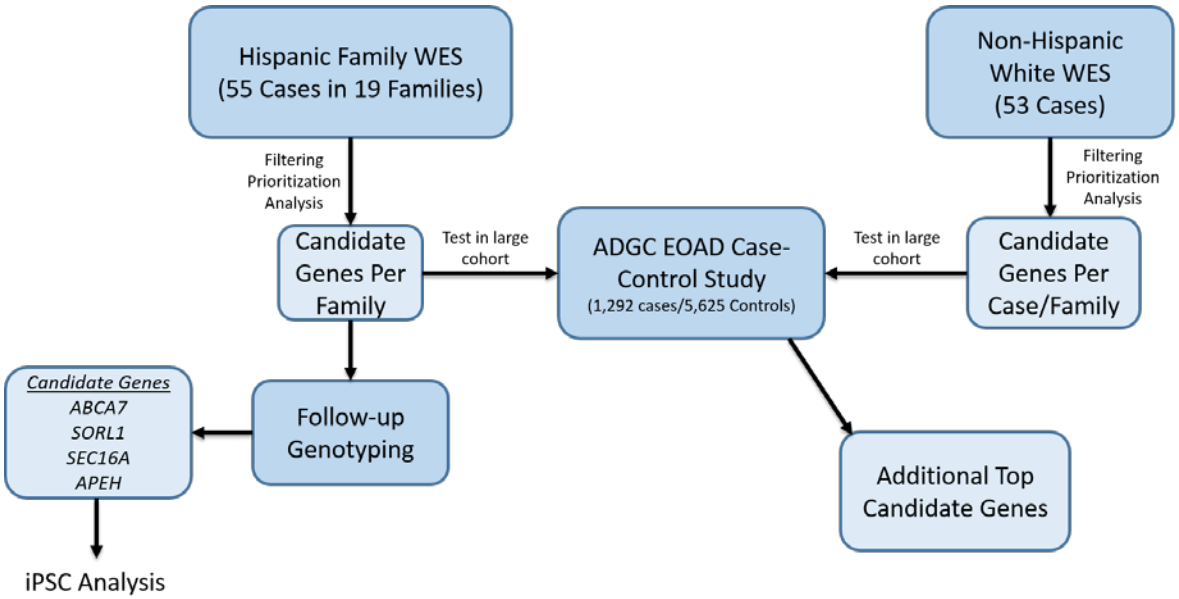

Figure 1. Analysis and prioritization of candidate variants/genes flowchart.

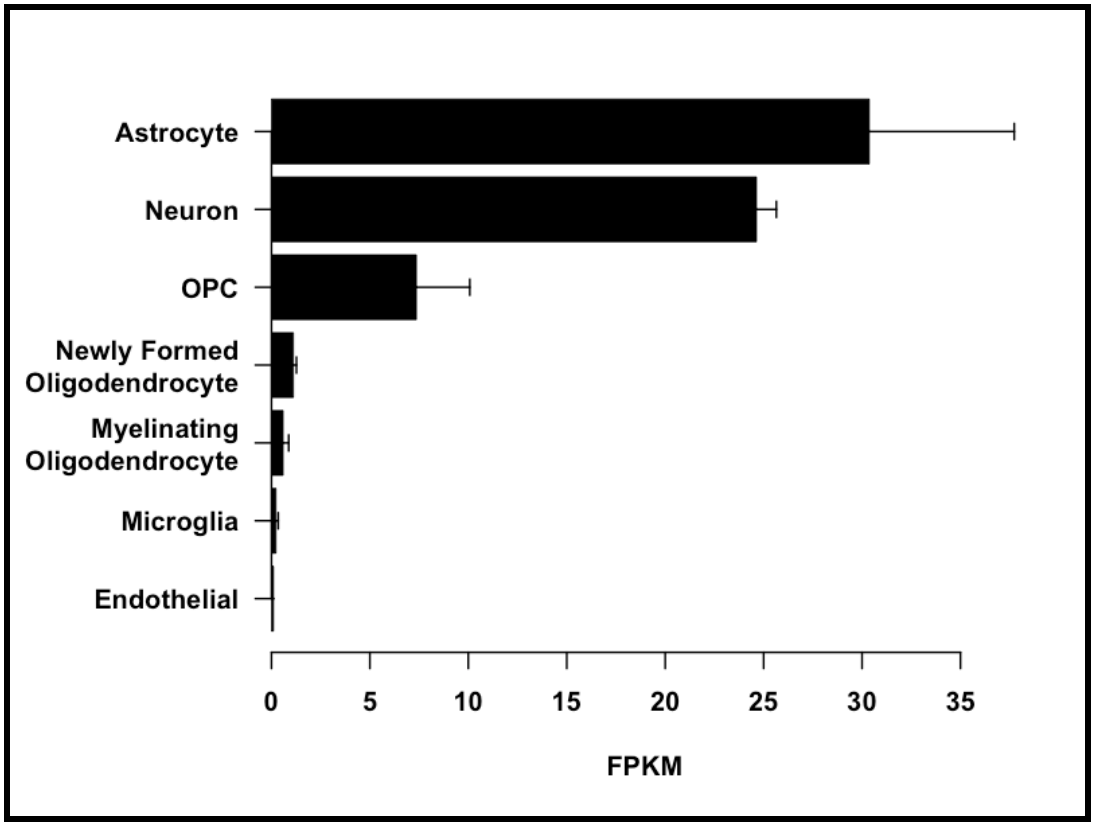

Figure 2. *PSD2* rna-seq levels in brain from the BRAIN-RNASeq database (Zhang et al. 2014).

Table 1. Rare (MAF<0.001), nonsynonymous or loss-of-function variants (LOF) found in two or more EOAD cases. Additional rare, missense or LOF variants found in these genes in other cases are also listed.

Chr:Position	Ref/Alt Allele	rsID	Gene	Consequence	N Affected	MAF (ExAC)	CADD	STRINGdB Interaction
1:9795622	C/T	.	CLSTN1	missense	2	8.95E-05	15.92	APP, PSEN1, OGT
3:51251601	G/A	rs199600118	DOCK3	missense	2	5.80E-04	29.5	PSEN1, PSEN2
3:51312575	G/C	.	DOCK3	missense	1	N	29	PSEN1, PSEN2
6:161969982	G/+C	.	PARK2	frameshift	1	8.13E-06	35	APP, MAPT, PSEN1
6:161990390	C/G	rs72480423	PARK2	missense	3	1.95E-04	22.4	APP, MAPT, PSEN1

Table 2. Top Results, ranked by P-value, of ADGC Exome Chip Analysis. Genes above the red line are genome-wide significant at a Bonferonni correction for 7249 genes tested (P=6.90 x 10-3); The blue line represents significance for 2+ non-Hispanic White (NHW) EOAD cases with a rare, damaging variant in the same gene (910 genes tested; P=5.49 x 10-5); The yellow line represent significance for 2+ Hispanic families with a segregating rare, damaging variant in the gene same gene (73 genes tested; 6.85 x 10-4)

Gene	P-Value	N Rare SNPs Tested	N NHW Cases*	N Hispanic Families**
RFTN1	7.35E-08	9		
PER3	1.74E-07	29		1
SH2B3	2.08E-07	13		
ZFYVE9	3.52E-07	25		
MYEOV2	4.86E-07	5		
C1GALT1	5.02E-07	3		
PSD2	6.98E-07	17	2	
PCDHB11	8.92E-07	10		1
BSG	1.02E-06	6		
PADI1	2.57E-05	16		
MUC17	3.70E-05	66		
TGFB1	5.31E-05	3		
PKD1	1.52E-04	18		
LONP1	1.61E-04	23		
NPC1L1	2.01E-04	31		
RBFOX1	2.38E-04	10		1
ABR	2.43E-04	6		
EXD3	4.61E-04	49		
KLHDC7B	4.73E-04	4		
P2RY4	5.47E-04	11	2	
MEGF8	6.04E-04	24	3	
EMID1	6.07E-04	5		
ADAM17	7.85E-04	15		
FBF1	8.18E-04	18	4	
PBLD	8.18E-04	7		
C20orf123	8.22E-04	5		
CEACAM20	8.29E-04	16		
IL16	8.33E-04	20	2	2
MAPK11	9.87E-04	2		

*Number of NHW cases with a rare, damaging variant in the gene

**Number of Hispanic families with a rare, damaging variant in the gene

iPSC-based functional characterization

During the current funding period, we were able to collect peripheral blood mononuclear cells (PBMCs) from the whole blood of AD individuals, as well as, race and gender-matched control individuals. These PBMCs were reprogrammed in iPSC through the transient overexpression of the Yamanaka factors – OCT4, SOX2, KLF2, and c-MYC using the Sendai virus system. We have derived multiple lines from non-hispanic white individuals bearing variants in the SORL1 or TTC3 gene, African American individuals bearing an ethnic-specific deletion in the ABCA7 gene, and Caribbean-Hispanic (Dominican Republic) individuals bearing variants in the SEC16A gene. These lines have been characterized for their pluripotency by immunocytochemistry (ICC), functional pluripotency through embryoid body formation, and karyotype analysis to ensure the stability of the genomes.

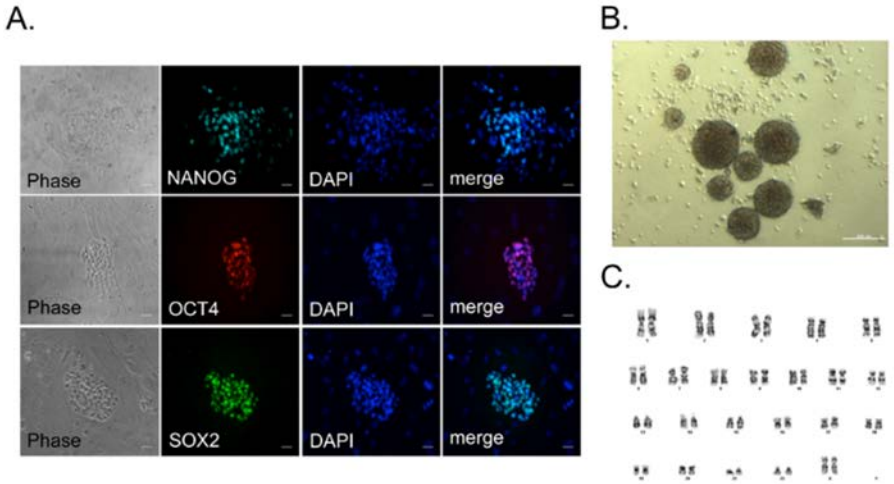

Figure X. **Figure 1. Validation of pluripotency of AD iPSC bearing mutations in the ABCA7 gene.** A. Immunocytochemistry for pluripotency factors (Nanog, Oct4 and Sox2). B) These iPSC have the capacity to form embryoid bodies (a test of pluripotency) and were found to have a normal karyotype (C).

We have begun the differentiation of the iPSC lines into forebrain neurons using a multistep approach beginning with the formation of neurospheres, the plating of the neurospheres on poly-L-Ornithine/laminin to form neural rosettes, the formation and expansion of neural progenitor cells. The neural progenitor cells will then be replated on poly-D-lysine/laminin and differentiated into forebrain neurons (as determined by staining for the expression of the appropriate markers). These neurons were analyzed at different time post initiation of differentiation to identify the optimal timing for the analysis of the different amyloid beta species (Aβ40 and Aβ42) from the culture supernatant and tau and phospho-tau species from intracellular lysates of the iPSC derived neurons. In our preliminary results, we found that there were elevated levels of Aβ40 in the ABCA7 deletion bearing sample.

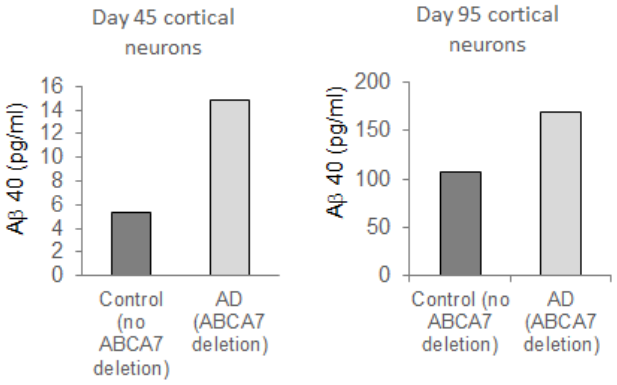

Figure X. **Amyloid beta 40 levels in iPSC-derived neurons at days 45 and 90 post initiation of differentiation in an AD-specific neurons (ABCA7 deletion bearing) compared to matched controls.**

We have also optimized the timing of the Aβ42 and tac/phosphor tau expression. We are currently differentiating the lines bearing mutations in the SORL1 gene, SEC16A, and TTC3 genes, as well as, matched control samples. All of these samples will be analyzed for the expression of amyloid beta species and pathogenic tau species, as well as, for

morphological changes in the cells and the rates of cell death/apoptosis.

REFERENCES

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014 Feb 2.

Zhang Y, Chen K, Sloan S, Bennett M, Scholze A, O'Keefe S, Phatnani H, Guarnieri P, Caneda P, Ruderisch N, Deng S, Liddelow S, Zhang C, Daneman R, Maniatis T, Barres B, Wu J. An RNA-Seq transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. Journal of Neuroscience. 2014.

KEY RESEARCH ACCOMPLISHMENTS:

- Publication of the link between a 44 base-pair deletion in the known LOAD gene *ABCA7* as a risk factor for Alzheimer's in both Hispanics and African Americans.
- Submittal of a manuscript describing a link between SORL1 and Parkinsonism in Alzheimer's cases.
- Sanger confirmation and genotyping in cases and controls of variants in 5 genes that interact with known EOAD genes and have shared rare coding variants with damaging potential in 2 or more NHW EOAD was completed. No significant variant was identified.
- Analysis of EOAD exome chip association data and comparison to WES sequencing identifies several candidate genes for EOAD, including the endocytosis related gene *PSD2*.
- Ascertainment of patient samples for iPSC derivation from the Hussman Institute Human Genomics (HIHG) cohort and the Columbia University cohort with genetic variants in SORL1, SEC16A, TTC3, and ABCA7.
- PBMCs reprogrammed into iPSC from the patient samples and validated for pluripotency and karyotype.
- Optimization of assay for pathogenic beta species from the culture supernatant and pathogenic tau from lysates from iPSC-derived neurons.

REPORTABLE OUTCOMES:

Platform Presentation at AAIC 2015 (Appendix I): Gary W. Beecham, PhD; Brian W. Kunkle, PhD, MPH; Badri Vardarajan, PhD; Patrice L. Whitehead, BS; Sophie Rolati, MS; Eden R. Martin, PhD; John R. Gilbert, PhD. Whole-Exome Sequencing in Early-Onset Alzheimer Disease Cases Identifies Novel Candidate Genes. The Annual Alzheimer's Association International Conference, Washington, D.C. July 18-23, 2015.

Poster Presentation at AAIC 2015 (Appendix II): Holly N. Cukier, PhD; Brian W. Kunkle, PhD, MPH; Sophie Rolati, MS; Kara L. Hamilton-Nelson, MPH; Martin A. Kohli, PhD; Beth A. Dombroski, PhD; Badri N. Vardarajan, PhD; Patrice L. Whitehead, BS; Derek J. Van Booven, BS; Eden R. Martin, PhD; Gary W. Beecham, PhD; Lindsay A. Farrer, PhD; Michael L. Cuccaro, PhD; Jeffery M. Vance, MD, PhD; Richard Mayeux, MD, MSc; John R. Gilbert, PhD; Regina M. Carney, MD; Goldie S. Byrd, PhD; Jonathan L. Haines, PhD; Gerald D. Schellenberg, PhD; Margaret A. Pericak-Vance, PhD; Rosalyn Lang, PhD and Alzheimer Disease Genetics Consortium. ABCA7 Deletion Associated with Alzheimer's Disease in African Americans. The Annual Alzheimer's Association International Conference, Washington, D.C. July 18-23, 2015.

Poster Presentation accepted at ASHG 2015 (Appendix III): Brian W. Kunkle, Badri Vardarajan, Patrice L. Whitehead, Sophie Rolati, Eden R. Martin, John R. Gilbert, Richard P. Mayeux, Jonathan L. Haines, Margaret A. Pericak-Vance, Gary W. Beecham. Whole-exome sequencing identifies novel candidate genes for early- onset Alzheimer disease. The 65th Annual Meeting of The American Society of Human Genetics, Baltimore, MD, October 6-10, 2015.

Poster Presentation accepted at ASHG 2015 (Appendix IV): Cuccaro ML, Carney RM, Kunkle BW, Vance JM, Whitehead PL, Gilbert JR, Vardarajan BN, Haines JL, Mayeux R, Pericak-Vance MA. SORL1 mutations and Parkinsonian features in early onset Alzheimer's disease families. The 65th Annual Meeting of The American Society of Human Genetics, Baltimore, MD, October 6-10, 2015.

Poster Presentation (Appendix V): Dykxhoorn DM, Cukier HN, Kunkle BW, Vardarajan BN, Rolati S, Hamilton-Nelson KL, Kohli MA, Whitehead PL, Van Booven DJ, Lang R, Farrer LA, Cuccaro ML, Vance JM, Gilbert JR, Beecham GW, Martin ER, Carney RM, Mayeux RP, Schellenberg GD, Byrd GS, Haines JL, Pericak-Vance MA, ADGC. ABCA7 Frameshift Deletion Associated with Alzheimer's Disease in African Americans. The Annual Alzheimer's Association International Conference, Toronto, Canada, July 24-28, 2016.

Platform Presentation (Appendix VI): Kunkle BW, Vardarajan BN, Naj AC, Cukier HN, Dyxkhoorn DM, Rolati S, Whitehead PL, Carney RM, Cuccaro ML, Vance JM, ADGC, Farrer LA, Haines JL, Schellenberg GD, Martin ER, Reitz C, Beecham GW, Mayeux R, Pericak-Vance MA. Identification of Novel Candidate Genes for Early-Onset Alzheimer Disease through Integrated Whole-Exome Sequencing and Exome Chip Array Association Analysis. The Annual Alzheimer's Association International Conference, Toronto, Canada, July 24-28, 2016.

Poster Presentation (Appendix VII): Pericak-Vance MA, Kunkle BW, Carney RM, Kohli MA, Naj AC, Hamilton KL, Whitehead PL, Cuccaro ML, Vance JM, Byrd G, Beecham GW, Gilbert JR, Haines JL, Martin ER. Targeted sequencing of Late-Onset Alzheimer Disease Loci Identifies Genomic Regions with Potential Functional Variants. The 13th Annual International Congress of Human Genetics, Kyoto, Japan, April 3-7, 2016.

Platform Presentation (Appendix VIII): Gary W. Beecham, PhD; Brian W. Kunkle, PhD, MPH; Badri Vardarajan, PhD; Patrice L. Whitehead, BS; Sophie Rolati, MS; Eden R. Martin, PhD; John R. Gilbert, PhD. Novel candidate genes for early-onset Alzheimer disease identified using whole-exome sequencing. The 13th Annual International Congress of Human Genetics, Kyoto, Japan, April 3-7, 2016.

Manuscripts

ABCA7 deletion associated with Alzheimer's disease in African Americans. *Accepted and published in Neurology: Genetics*

Whole-exome sequencing of Hispanic families identifies novel candidate genes for early-onset Alzheimer disease. *In preparation.*

Whole-exome sequencing identifies novel candidate genes for early-onset Alzheimer disease. *In preparation.*

SORL1 mutations and Parkinsonian features in early onset Alzheimer's disease families. *Submitted to Alzheimer's and Dementia*

CONCLUSION:

Mutations in *APP*, *PSEN1* and *PSEN2* lead to familial EOAD and accounting for 60-70% of familial EOAD and ~11% of EOAD overall, leaving the majority of genetic risk for this form of Alzheimer disease unexplained. We performed Whole-Exome Sequencing (WES) on 55 individuals in 19 Caribbean Hispanic EOAD families and 51 Non-Hispanic White EOAD cases previously screened negative for *APP*, *PSEN1* and *PSEN2* to search for rare variants contributing to risk for EOAD. Variants were filtered for segregating, conserved and functional rare variants (MAF<0.1%) assuming both autosomal and X-linked dominant models. We have identified and published or submitted manuscripts on variants identified in these analyses. We continue to identify candidate risk genes for EOAD, including an endocytic gene, *PSD2*, which we find to be significant in an analysis of EOAD exome chip association data from the ADGC. A comparison of these results to our EOAD WES sequencing identified two NHW cases with rare, damaging, missense variants in the *PSD2* gene. We also are following up our most promising results in iPSC analysis and have ascertained patient samples from the Hussman Institute Human Genomics (HIHG) cohort and the Columbia University cohort with genetic variants in SORL1, SEC16A, TTC3, and ABCA7 for these analyses. PBMCs have been reprogrammed into iPSC from the patient samples and validated for pluripotency and karyotype, and optimization of the assay for pathogenic beta species from the culture supernatant and pathogenic tau from lysates from iPSC-derived neurons is complete.

APPENDICES:

Appendix I:

Whole-Exome Sequencing in Early-Onset Alzheimer Disease Cases Identifies Novel Candidate Genes

Gary W. Beecham, PhD[1]; Brian W. Kunkle, PhD, MPH[1]; Badri Vardarajan, PhD[2]; Patrice L. Whitehead, BS[1]; Sophie Rolati, MS[1]; Eden R. Martin, PhD[1]; John R. Gilbert, PhD[1]; Richard Mayeux, MD, MSc[2]; Jonathan L. Haines, PhD[3] and Margaret A. Pericak-Vance, PhD[1], (1)University of Miami, Miami, FL, USA, (2)Columbia University, New York, NY, USA, (3)Case Western Reserve University, Cleveland, OH, USA

Abstract

Background: Mutations in *APP*, *PSEN1* and *PSEN2* lead to early-onset Alzheimer disease (EOAD). These mutations account for ~11% of EOAD overall, leaving the majority of genetic risk for the most severe form of Alzheimer disease unexplained.

Methods: We performed Whole-Exome Sequencing (WES) in 50 Caucasian EOAD cases previously screened negative for *APP*,*PSEN1*, and *PSEN2* to search for rare variants contributing to risk for EOAD. Variant filtering for functional, damaging rare variants (MAF<0.1%) was performed. Genes with shared (2+ cases with the same variant), damaging variants were examined for interactions with known EOAD genes (*APP, PSEN1, PSEN2, SORL1, GRN*, *MAPT*) and *APOE* using STRINGdb.

Results: 176 genes had rare functional variants shared in two or more cases. 46 of these genes were prioritized for their damaging potential, defined by their shared rare variants having a Combined Annotation Dependent Depletion (CADD) score in the top 10% of all variants. Gene network analysis of these 46 genes with known EOAD genes and *APOE* identified five top candidate genes: *HSPG2* (interacts with*GRN*, *APOE*, and *APP*), *CLSTN1* (interacts with *PSEN1* and *APP*), *DOCK3* (interacts with *PSEN1* and *PSEN2*), and the *APOE*interactors *SAR1B* and *STAT1*. 5 cases have a variant in *HSPG2*, a gene potentially involved in amyloidogenesis and tau aggregation in AD, while 4 cases have a variant in *DOCK3*, a gene expressed exclusively in the central nervous system and associated with neurofibrillary tangles in AD brains.

Conclusions: WES of EOAD cases identified several genes with potential roles in AD pathogenesis.

Appendix II:

ABCA7 Deletion Associated with Alzheimer's Disease in African Americans

Holly N. Cukier, PhD[1]; Brian W. Kunkle, PhD, MPH[1]; Sophie Rolati, MS[1]; Kara L. Hamilton-Nelson, MPH[1]; Martin A. Kohli, PhD[1]; Beth A. Dombroski, PhD[2]; Badri N. Vardarajan, PhD[3]; Patrice L. Whitehead, BS[1]; Derek
J. Van Booven, BS[1]; Eden R. Martin, PhD[1]; Gary W. Beecham, PhD[1]; Lindsay A. Farrer, PhD[4]; Michael L. Cuccaro, PhD[1]; Jeffery M. Vance, MD, PhD[1]; Richard Mayeux, MD, MSc[3]; John R. Gilbert, PhD[1]; Regina M. Carney, MD[1]; Goldie S. Byrd, PhD[5]; Jonathan L. Haines, PhD[6]; Gerald D. Schellenberg, PhD[2]; Margaret A. Pericak-Vance, PhD[1]; Rosalyn Lang, PhD[5] and Alzheimer Disease Genetics Consortium, (1)University of Miami, Miami, FL, USA, (2)University of Pennsylvania, Philadelphia, PA, USA, (3)Columbia University, New York, NY, USA, (4)Boston University, Boston, MA, USA, (5)North Carolina A&T State University, Greensboro, NC, USA, (6)Case Western Reserve University, Cleveland, OH, USA

Abstract

Background: The *ATP-binding cassette, sub-family A (ABC1), member 7* (*ABCA7*) gene has been implicated as a risk factor in Alzheimer's disease (AD) in both African American and Caucasian populations. However, the effect in African Americans is significantly higher, comparable to that found in *APOE ϵ4*. Furthermore, the underlying damaging allele(s) conveying the strong genome-wide signal has yet to be revealed.

Methods: We performed custom next generation sequencing across the *ABCA7* region on 41 African American individuals with AD and 48 control African Americans carrying the previously reported risk allele, rs115550680. Using Agilent custom capture, probes were designed across a 150 kb genomic area that includes*ABCA7*as well as eight flanking genes and a small nuclear RNA. Samples were run on the Illumina HiSeq sequencing machine. Data processing was performed with Casava, GATK, and BWA, and deletions were identified with Pindel. Our top priority alterations were confirmed by Sanger sequencing and validated in two African American cohorts – HIHG (cases: 482, controls: 565) and ADGC (cases: 746, controls: 1,063). Analyses for each cohort were adjusted for gender, and age (controls >65).

Results: 1,120 SNVs were detected by sequencing. 11 variants had different frequencies in cases and controls ($p<0.1$). In addition, a 44 base pair deletion (rs142076058) was identified in all 89 cases and controls, signifying that it could be in high linkage disequilibrium with the risk allele. The deletion was significantly correlated with disease in the HIHG cohort ($p = 0.020$, odds ratio = 1.54, 95% confidence interval 1.07-2.22). This finding was then replicated in the ADGC cohort ($p = 0.006$, odds ratio = 1.50, 95% confidence interval 1.12-2.01). The deletion falls in the 14th exon of*ABCA7*and results in a frameshift and truncating mutation (p.Arg578Alafs) that could interfere with protein function. When we investigated this same deletion in our cohort of European ancestry (>3,000 samples), the deletion was only found in 5 individuals, 4 of whom also carried the risk allele, and did not correlate with disease.

Conclusions: This deletion in *ABCA7* could represent an ethnically-specific pathogenic alteration in Alzheimer's disease.

Appendix III:

Whole-exome sequencing identifies novel candidate genes for early-onset Alzheimer disease

Brian W. Kunkle, Badri Vardarajan[2], Patrice L. Whitehead[1], Sophie Rolati[1], Eden R. Martin[1], John R. Gilbert[1], Richard P. Mayeux[2], Jonathan L. Haines[3]. Margaret A. Pericak-Vance[1], Gary W. Beecham[1].

[1]John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA
[2]Taub Institute of Research on Alzheimer's Disease, Columbia University, New York, NY, USA
[3]Institute for Computational Biology, Case Western Reserve University, Cleveland, OH, USA

Objectives

Mutations in *APP*, *PSEN1* and *PSEN2* lead to early-onset Alzheimer disease (EOAD). These mutations account for ~11% of EOAD overall, leaving the majority of genetic risk for the most severe form of Alzheimer disease unexplained.

Methods

To search for rare variants contributing to risk for EOAD we performed Whole-Exome Sequencing (WES) in 50 Caucasian EOAD cases screened negative for *APP*, *PSEN1*, and *PSEN2*. Variant filtering for functional, damaging rare variants (MAF<0.1%) was performed. Damage prediction was performed using a CADD score (Combined Annotation Dependent Depletion; top 10%). Rare, damaging variants shared by multiple cases (+2) were then selected for follow-up protein-protein interaction analysis with known EOAD genes (*APP, PSEN1, PSEN2, GRN*, *MAPT*) using the program STRINGdb. Additionally, we performed a separate text-mining screen of this set of variants using Phenolyzer and the terms 'Alzheimer disease', 'dementia', and 'neurodegeneration'.

Results

We identified 51 rare, damaging variants in 46 genes in two or more EOAD cases. Gene network analysis of these 46 genes with known EOAD genes identified five candidate genes: HSPG2 (interacts with GRN and APP), CLSTN1 (interacts with PSEN1 and APP), and DOCK3 (interacts with PSEN1 and PSEN2). Five cases have a variant in HSPG2, a gene in a LOAD susceptibility region and potentially involved in amyloidogenesis and tau aggregation in AD. Four cases have a variant in DOCK3, a gene shown to regulate amyloid-β secretion, and associated with neurofibrillary tangles in AD brains. Several other cases have shared variants in CLSTN1. Disruption of calsyntenin-1-associated axonal transport of APP by mutations in CLSTN1, a known APP interactor, have been identified as a potential pathogenic mechanism of Alzheimer's. Moreover, CLSTN1's potential as a regulator of synapse formation and neuronal development suggests other mechanisms through which it could be involved in development of dementia. Top scoring genes for the Phenolyzer analysis included the genes HSPG2, STAT1, a BACE1 interactor, and CNTNAP2, whose expression is down-regulated in AD patients.

Conclusions

WES of EOAD cases identified several genes with potential roles in AD pathogenesis.

Appendix IV:

SORL1 mutations and Parkinsonian features in early onset Alzheimer's disease families

Cuccaro ML, Carney RM, Kunkle BW, Vance JM, Whitehead PL, Gilbert JR, Vardarajan BN, Haines JL, Mayeux R, Pericak-Vance MA

Early-onset Alzheimer's disease (EOAD; onset of Alzheimer's disease (AD) prior to age 65) affects ~200,000 individuals, representing 3.7% of all AD cases. Mutations in *PSEN1*, *PSEN2*, and *APP* can cause EOAD. We conducted a genome-wide search to identify novel mutations in a series of families enriched for EOAD using whole-exome sequencing (WES). We hypothesized that WES would identify novel mutations in these families and that these mutations would be associated with a spectrum of neurodegenerative features.

Our dataset consisted of 50 families with at least one individual with EOAD. All individuals and additional family members were characterized via comprehensive medical and research record review. WES (Agilent SureSelect) was performed on individuals of interest per family; *APOE* genotypes were available as well.

WES identified variants of interest in *SORL1*, a gene which previously had been primarily associated with late-onset AD. *SORL1* missense mutations were identified in 2 families. One family contained 4 affected individuals (AOO range 59-82 years; 3 with APOE genotype 3/4, 1 with 3/3) and 2 unaffected individuals (ages 81-84, both APOE 3/3) with the novel T588I *SORL1* mutation. Two individuals with *SORL1* mutations in this family also presented with Parkinsonian features. The second family carried the previously reported *SORL1* mutation (T749M) with 3 affected individuals with the mutation (AOO 55-84, all APOE 3/3), 1 affected individual without the mutation (AOO 76, APOE 3/4), and 1 unaffected individual with the mutation (age 79, APOE 3/3). One affected individual with in the second family with the T749M *SORL1* mutation had neuropathologic evidence of Lewy bodies without clinical Parkinsonism.

A follow-up of SORL1 mutations in a separate dataset of LOAD families (Vardarajan 2015, Annals of Neurology) corroborated the presence of Parkinsonism in 4 individuals with *SORL1* mutations (3 individuals with the common rs2298813 mutation and 1 individual with a rare mutation).

This study confirms *SORL1* as a gene involved in risk for EOAD. Equally compelling is our finding of Parkinson-related features and Lewy bodies without clinical Parkinsonism in association with these *SORL1* mutations. This is the first study to suggest a relationship between *SORL1*, AD, and Parkinsonism.

Appendix V:

ABCA7 Frameshift Deletion Associated with Alzheimer's Disease in African Americans

Derek M. Dykxhoorn, Holly N. Cukier, Brian W. Kunkle, Badri N. Vardarajan, Sophie Rolati, Kara L. Hamilton-Nelson, Martin A. Kohli, Patrice L. Whitehead, Derek J. Van Booven, Rosalyn Lang, Lindsay A. Farrer, Michael L. Cuccaro, Jeffery M. Vance, John R. Gilbert, Gary W. Beecham, Eden R. Martin, Regina M. Carney, Richard P. Mayeux, Gerald D. Schellenberg, Goldie S. Byrd, Jonathan L. Haines, Margaret A. Pericak-Vance, and the Alzheimer Disease Genetics Consortium.

Background: The *ATP-binding cassette, sub-family A (ABC1), member 7* (*ABCA7*) gene has been implicated as a risk factor in Alzheimer's disease (AD) in African American (AA), Asian, and non-Hispanic white (NHW) populations. However, the effect in African Americans is significantly stronger, comparable to that found in *APOE ϵ4*. While some rare loss-of-function variants in *ABCA7* were identified in NHW populations that may contribute to AD pathogenicity, potentially pathogenic variants have yet to be reported in AA populations.

Methods: We performed custom next generation sequencing across the *ABCA7* region on 40 AA individuals with AD and 37 control AA carrying the previously reported risk allele, rs115550680. Custom capture was performed across a 150 kb genomic area encompassing *ABCA7* and sequenced on the Illumina Hiseq. Data processing was performed with Casava, GATK, and BWA, and deletions were identified with Pindel. Our top variants were confirmed by Sanger sequencing and validated in two AA cohorts – HIHG (cases: 539, controls: 529) and ADGC (cases: 687, controls: 1,062). Whole exome sequencing was performed on 19 Caribbean Hispanic multiplex AD families (46 cases and 6 unaffected relatives). iPSC lines were developed from AA AD individuals bearing the *ABCA7* deletion which were differentiated into cortical neurons and functionally characterized.

Results: A 44 base pair deletion (rs142076058) was identified in all 77 risk genotype carriers. The deletion was significantly associated with disease (p=0.0002, OR=2.13 [95% CI:1.42-3.20]) in the HIHG cohort and replicated in the ADGC cohort ($p=1.414x10^{-5}$, OR=1.81 [95% CI:1.38-2.37]). The deletion is common in AA cases (15.2%) and AA controls (9.74%), but in only 0.12% of our NHW cohort. Whole exome sequencing of multiplex, Caribbean Hispanic families identified the deletion co-segregating with disease in a large sibship. The deleted allele produces a stable, detectable RNA strand and is predicted to result in a frameshift mutation (p.Arg578Alafs) that could interfere with protein function. IPSC-derived cortical neurons bearing the *ABCA7* deletion produced elevated levels of pathogenic □-amyloid production compared to controls.

Conclusions: The deletion in *ABCA7* could represent an ethnically-specific pathogenic alteration in Alzheimer's disease that results in impaired APP processing and increased toxic □-amyloid production.

Theme: Genetics, Genetic factors of Alzheimer's disease

Appendix VI:

Identification of Novel Candidate Genes for Early-Onset Alzheimer Disease through Integrated Whole-Exome Sequencing and Exome Chip Array Association Analysis

Brian W. Kunkle, PhD, MPH[1], Badri N. Vardarajan, PhD[2], Adam C. Naj, PhD[3], Holly N. Cukier, PhD[1], Derek M Dykxhoorn, Ph.D.[1], Sophie Rolati, MS[1], Patrice L. Whitehead, BS[1], Regina M. Carney, MD[1], Michael L. Cuccaro, PhD[1], Jeffery M. Vance, MD, PhD[1], Alzheimer's Disease Genetics Consortium[4], Lindsay A. Farrer, PhD[5], Jonathan L. Haines, PhD[6], Gerard D Schellenberg, PhD[3], Eden R. Martin, PhD[1], Christiane Reitz, MD PhD[2], Gary W. Beecham, PhD[1], Richard Mayeux, MD, MSc[2]and Margaret A. Pericak-Vance, PhD[1], (1)University of Miami, Miami, FL, USA, (2)Columbia University, New York, NY, USA, (3)University of Pennsylvania, Philadelphia, PA, USA, (4)University of Pennsylvania School of Medicine, Philadelphia, PA, USA, (5)Boston University, Boston, MA, USA, (6)Case Western Reserve University, Cleveland, OH, USA

Abstract Text:
Background: Mutations in APP, PSEN1 and PSEN2lead to familial early-onset Alzheimer disease (EOAD). These mutations account for ~11% of EOAD overall, leaving the majority of genetic risk for the most severe form of AD unexplained. Methods: We performed whole-exome sequencing (WES) on 53 Non-Hispanic White EOAD cases screened negative for APP, PSEN1, and PSEN2 to search for rare EOAD risk variants. Variant filtering for missense and loss-of-function (LOF) rare variants (MAF<0.1%) was performed, and filtered variants present on the Illumina exome chip array were tested in a cohort of 1,292 EOAD cases (Age-at-onset<65) and 5,625 controls (Age≥65) from the Alzheimer's Disease Genetics Consortium (ADGC). As no LOF filtered variants were on the chip, we assessed these variants by using both the lowest variant P-value per gene and gene-based SKAT-O results. Rare LOF variants were prioritized for testing if they were in 2+ cases and in haploinsufficient genes (Haploinsufficient Score≥50). Results: 1,803 rare missense variants identified in the WES sample were available for testing in in the exome chip study. 70 of these variants were nominally associated with EOAD, with the strongest signals in the neuropeptide NPPB (OR=11.24, P=0.003), and NDST2 (OR=14.95, P=0.001), a processor of heparin sulfate, a molecule with potential importance in Aβ formation. Assessment of the 70 nominally significant genes containing these variants for consistent dysregulation (3+ expression studies) in AD using AlzBase prioritized 11 genes with strong potential for involvement in EOAD, including the APP interector ADRA1A (OR=3.60, P=0.032). One variant in the LOAD GWAS gene RIN3 (OR=6.94, P=0.009) also was of interest. Testing of 30 LOF genes revealed several associations that approached significance including a splicing variant in CACNA1G (OR=4.93, P=0.002), a gene associated with cognitive decline and age-related production of Aβ, a stopgain in CENPF (OR=11.46, P=0.004), a gene involved in endolysosomal transport and amyloid plaque generation, and a frameshift in the gene TDP2 (Gene P=0.006), for which homozygous LOF mutations cause neurodegeneration with epilepsy. Conclusions: Testing of candidate WES EOAD risk variants and genes in the ADGC EOAD exome chip study identified several genes with potential roles in AD pathogenesis.

Word Limit: 350 **Word Count: 345**

Appendix VII:

Targeted sequencing of Late-Onset Alzheimer Disease Loci Identifies Genomic Regions with Potential Functional Variants

M.A. Pericak-Vance[1], B. W. Kunkle[1], R. M. Carney[2], M. A. Kohli[1], A.C. Naj[1,3], K. L. Hamilton[1], P. L. Whitehead[1], L. Wang[1], M. L. Cuccaro[1], J. M. Vance[1], G. Byrd[4], G. W. Beecham[1], J. R. Gilbert[1], J. L. Haines[5], E. R. Martin[1]

1) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 3) Department of Biostatistics & Epidemiology, University of Pennsylvania, Philadelphia, PA 2) Department of Psychiatry & Behavioral Sciences, University of Miami, Miami, FL; 4) Department of Biology, North Carolina A & T State, Greensboro, NC 5) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH.

ABSTRACT

BACKGROUND: Genome-wide association studies (GWAS) have identified multiple loci that confer late-onset Alzheimer's disease (LOAD) risk. Few causal variants and genes at these loci have been identified however. We conducted targeted sequencing of nine of these loci (*ABCA7, BIN1, CD2AP, CD33, CLU, CR1, EPHA1, MS4A4A/MS4A6A*, and *PICALM*) in 291 Non-Hispanic White (NHW) LOAD cases and 103 normal controls in order to identify potential functional variants for follow-up experiments.

METHODS: We prioritized variants on three criteria: (a) functionality (missense, nonsense, or splice-site variant), (b) damaging potential, and (c) overrepresentation in cases vs. controls. We also examined variants grouped by gene and function using multilocus association testing. Follow-up genotyping of 1) rare, damaging variants prioritized by filtering and 2) 10 recently associated variants in these loci, was conducted in 2468 NHW familial and sporadic cases and controls and 892 African-American (AA) cases and controls.

RESULTS: Follow-up genotyping of 95 functional, damaging variants enriched or present only in cases identified a 3-UTR splice variant in *ABCA7* (chr19:1054190) present only in affecteds (N=5), including all 4 affected siblings of one family. Additionally, follow-up genotyping of previously reported rare variants in known AD loci confirmed the association of an intronic SNP (rs78117248) in the *ABCA7* gene in NHW Americans, a result originally reported in a Belgian population (P=0.0002). No significant results were found in the AA cohort. Region-based analysis of the sequencing results identified several clusters of variants associated with LOAD including seven near *MS4A* (P=0.0049) and seven upstream of *BIN1* (P=0.0067).

CONCLUSIONS: Targeted sequencing of LOAD risk loci identified rare and novel coding variants in LOAD susceptibility loci, suggesting that rare variants at these loci may contribute to LOAD risk.

CHARACTERS WITH SPACES: 1,933

Appendix VIII:

Novel candidate genes for early-onset Alzheimer disease identified using whole-exome sequencing

G. W. Beecham[1], B. W. Kunkle[1], B. Vardarajan[2], P. L. Whitehead[1], S. Rolati[1], E. R. Martin[1], J. R. Gilbert[1], R. P. Mayeux[1], J. L. Haines[3], M. A. Pericak-Vance[1]

1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Taub Institute of Research on Alzheimer's Disease, Columbia University, New York, NY, USA; 3) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH, USA.

BACKGROUND: Known mutations in *APP*, *PSEN1* and *PSEN2* account for ~11% of early onset Alzheimer's disease (EOAD), leaving the majority of EOAD's genetic risk unexplained.

METHODS: To search for novel variants contributing to EOAD risk we performed whole-exome sequencing (WES) in 50 Caucasian EOAD cases screened negative for *APP*, *PSEN1*, and *PSEN2*. Variants were filtered for functional, damaging rare variants (MAF<0.1%), as well as genotype quality (GQ>20), read depth (>5), and VQSLOD Score (>0). Variants were then filtered for nonsynonymous or splice site variants present in 2+ individual cases, and having a Combined Annotation Dependent Depletion (CADD) score in the top 10% of all variants. Surviving variants were assessed for protein-protein interaction with known EOAD genes (*APP, PSEN1, PSEN2, GRN*, *MAPT*) using STRINGdb.

RESULTS: We identified 51 rare, damaging variants in 46 genes in two or more EOAD cases. Gene network analysis of these 46 genes with known EOAD genes identified three candidate genes: HSPG2 (interacts with GRN and APP), CLSTN1 (interacts with PSEN1 and APP), and DOCK3 (interacts with PSEN1 and PSEN2). Specifically, five cases (2 pairs of cases with a shared variant and a fifth case with a different variant) have rare, damaging variation in *HSPG2,* a gene in a known AD susceptibility region that is potentially involved in amyloidogenesis and tau aggregation. Four cases (2 pairs of cases with a shared variant) have a variant in *DOCK3*, a gene shown to regulate amyloid-β secretion, and associated with neurofibrillary tangles in AD brains. Several other cases have shared variants in CLSTN1. Disruption of calsyntenin-1-associated axonal transport of APP by mutations in CLSTN1, a known APP interactor, has been identified as a potential pathogenic mechanism of AD.

CONCLUSION: This studiy identified three novel loci linked to EOAD risk. These genes are being investigated in additional datasets for further association with AD.